Hongdao Meng

mycrofthd@gmail.com | github.com/Mycroft-s | linkedin.com/in/hongdao-meng-70222b306 | (718) 3063737

EDUCATION

New York University Sep. 2024 - May 2026 (Expected)
Master of Science in Computer Science | GPA: 3.83 New York, NY

Beijing University of Technology Sep. 2020 - Jul. 2024
Bachelor of Engineering in Information Security | GPA: 4.00 Beijing, CN

SKILLS

Languages: C/C++, Java, Python, Go, SQL, JavaScript, HTML/CSS, Shell, PHP, LaTeX
Machine Learning: PyTorch, LangChain, Tensorflow, Pandas, Scikit-Learn, VGG16, ViT, Multi-View, Multi-Label
Frameworks: Blade, React, Angular, Vue.js, Django, Flask, Node.js, Spring Boot, Hadoop, Spark
Database: MySQL, Redis, MongoDB, PostgreSQL, DynamoDB, Oracle, Firebase, RocketMQ, Elasticsearch, MilvusDB
Tools: Git, Docker, AWS, Azure, CMake, Postman, CI/CD, Jenkins, Nginx, FFmpeg, OpenCV, Jira, Figma

PROFESSIONAL & RESEARCH EXPERIENCE

***Software Engineer Intern* @ TikTok, Recommendation System Infra, Seattle** May 2025 - Sep. 2025

- Developed a Python-based **DAG-DSL** migration toolkit (leveraging Protobuf and regex) to auto-convert legacy configuration files of the Bytedance Feature Service (BFS), a scalable feature extraction service powering TikTok's recommendation system, into standardized DSL definitions, boosting operator migration throughput by 85%.
- Designed and implemented a **Blade-based** C++ operator migration pipeline for BFS, streamlining the workflow to improve operator migration speed by 83.3%, and personally migrated 20 operators.
- Architected a one-stop workflow for User Data Accessor covering service build, auto generate python operators, DSL compilation, and RPC request, enabling collaboration among dozens of engineers and accelerating development by 73.3%.

***Machine Learning Engineer Intern* @ C2SMARTER Center, New York** Feb. 2025 - Present

- Led 6-member team to develop RAG-based chatbot system using **LangChain** and Flask-React, achieving **24.3%** accuracy improvement on MS MARCO dataset (F1=0.86) with **33.7%** faster response latency through query optimization
- Built hybrid retrieval framework with **MilvusDB** vector database and **BGE-M3** embeddings, improving search relevance by **21.7%** and boosting query performance **25.6%** through unified re-ranking architecture
- Implemented a Docker-based data pipeline with MongoDB on AWS EC2, reducing deployment setup time by 15.7% and deploying CI/CD pipelines with Jenkins to ensure high availability.

***Machine Learning Engineer and Founder* @ DeepFake Detection Startup, New York** Sep. 2024 - Dec. 2024

- Led 5-member team to develop core modules of deepfake detection web platform using React and TypeScript for seamless real-time interaction, enabling 1,200+ concurrent users and reducing client-side rendering latency by **21.3%**
- Fine-tuned **Vision Transformer** and **VGG16** models from **Hugging Face** for image and audio DeepFake detection, achieving **91.2%** and **88.1%** accuracy, respectively. Constructed two custom datasets from the MP4 FaceForensics corpus by extracting video frames with OpenCV and audio files with FFmpeg, and deployed both models on AWS EC2
- Built a real-time communication layer using Django and WebSocket for robust middleware communication, reducing task completion time by **25.6%** with <180ms P95 latency, and deployed backend services on Kubernetes with AWS ELB load balancing and HPA policies, achieving **99.5%** availability under 5k RPM
- Optimized **PostgreSQL** query execution through composite index tuning, reducing average response time by 18% (320ms→262ms)

***Machine Learning Engineer Intern* @ Data Mining & Security Lab, Beijing** Sep. 2022 - Jul. 2024

- Led research on Federated Multi-View Multi-Label Machine Learning. Published first-author paper in *IEEE Transactions on Big Data* 2025: "**Federated Multi-View Multi-Label Classification**" (DOI: 10.1109/TBDATA.2024.3522812)
- Devised solutions to complex data privacy challenges by proposing and developing the **FMVML** framework, a federated learning method enabling cross-view feature fusion and multi-label semantic classification, which outperformed all state-of-the-art methods, improving Average Precision by **8.3%** and lowering One Error by **14%**
- Utilized **Python**/**PyTorch** for model development and **Matlab** for signal processing; implemented data pipelines with **Pandas**/**Scikit-Learn**; produced publication-ready documents with LaTeX

PUBLICATION

- First Author: "Federated Multi-View Multi-Label Classification." *IEEE Transactions on Big Data*, 2025.
- Co-Author: "Susceptibility genes identification and risk evaluation model construction by transcriptome-wide association analysis for salt sensitivity of blood pressure: the EpiSS study." *BMC Genomics*, 2024.